Homework Assignment 3

Rushabh Ashok Dharia
CSCI-B555 Machine Learning

April 7, 2019

Note: Homework was discussed with Bhushan Malgaonkar

Ans 1. *a).*

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2}(\frac{y-Xw}{\sigma})^2}$$

$\square$

b).

$$\begin{aligned} f(w, y) &= f(y/w).f(w) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2}(\frac{y-Xw}{\sigma})^2} \frac{1}{\sqrt{2\pi\rho^2}} e^{\frac{-1}{2}(\frac{w}{\rho})^2} \\ &= \frac{1}{2\pi\sigma\rho} e^{\frac{-1}{2}[(\frac{y-Xw}{\sigma})^2+(\frac{w}{\rho})^2]} \end{aligned}$$

$\square$

c).

$$
\begin{aligned}
\hat{w}_{MAP} &= argmax(\frac{1}{2\pi\sigma\rho}e^{\frac{-1}{2}[(\frac{y-Xw}{\sigma})^2+(\frac{w}{\rho})^2]})\\
&\text{taking log}\\
&= argmax(-log(2\pi\sigma\rho) - \frac{1}{2}[(\frac{y-Xw}{\sigma})^2 + (\frac{w}{\rho})^2])\\
&= argmin(log(2\pi\sigma\rho) + \frac{1}{2}[(\frac{y-Xw}{\sigma})^2 + (\frac{w}{\rho})^2])\\
&log(2\pi\sigma\rho) \text{ is constant so we can neglect it}\\
&= argmin(\frac{1}{2}[(\frac{y-Xw}{\sigma})^2 + (\frac{w}{\rho})^2])\\
&\text{We can neglect } \frac{1}{2} \text{ as well}\\
&= argmin(\frac{y-Xw}{\sigma})^2 + (\frac{w}{\rho})^2)\\
&= argmin(\frac{(y-Xw)^T(y-Xw)}{\sigma^2} + \frac{(w)^T(w)}{\rho^2}\\
&||x|| = \sqrt{x^Tx}. \text{ Therefore, } ||x||^2 = x^Tx\\
&= argmin(\frac{||y-Xw||^2}{\sigma^2}) + (\frac{||w||^2}{\rho^2}))
\end{aligned}
$$

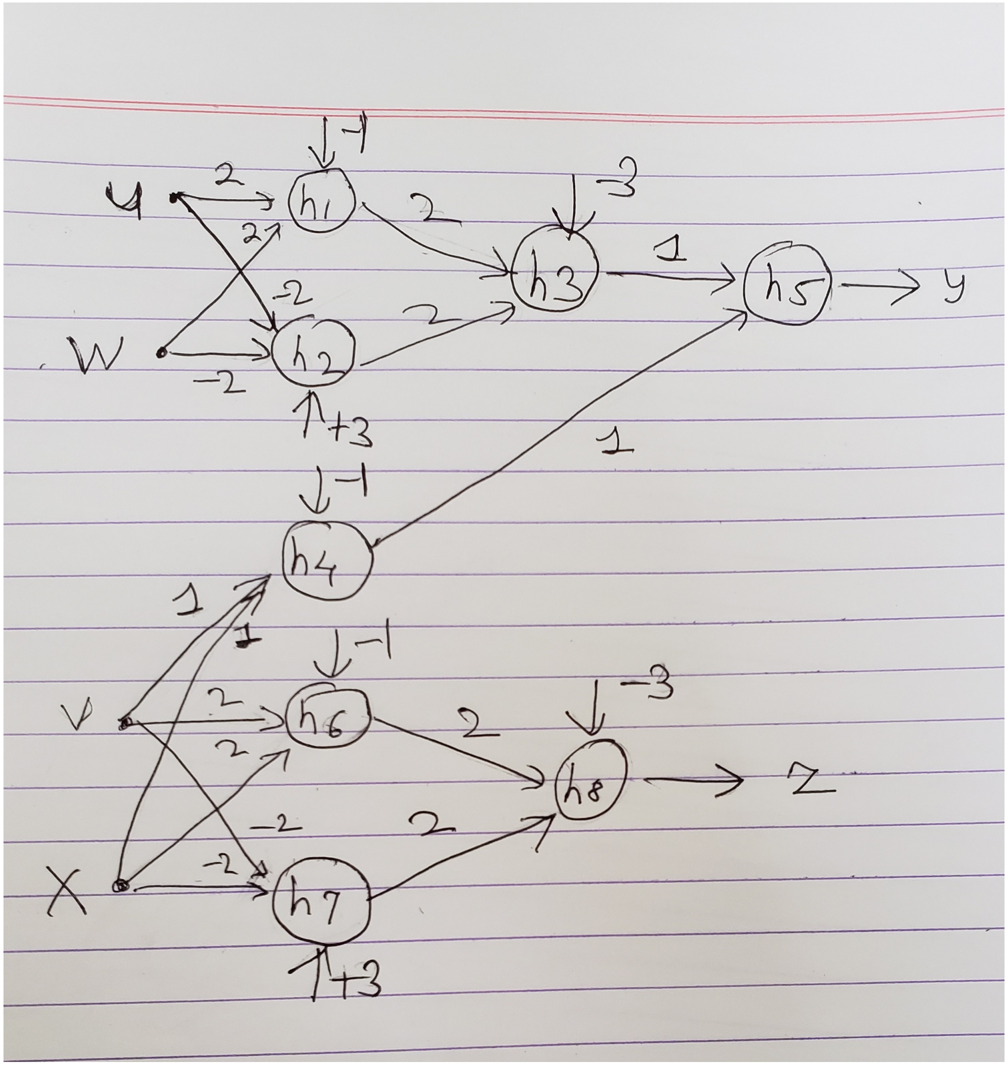
□

Ans 2.

$$
\begin{aligned}
y &= Xw + e\\
&= \mathcal{N}(Xw + \mu, \sigma^2)\\
\hat{y}_{MLE} &= argmax(\Pi\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-1}{2}[(\frac{y-Xw-\mu}{\sigma})^2}\\
&\text{taking log}\\
&= \sum -\frac{1}{2}log(2\pi\sigma^2) - \frac{1}{2}[(\frac{y-Xw-\mu}{\sigma})^2\\
&\text{taking derivative w.r.t } \mu\\
0 &= 0 + \frac{1}{2}\sum[(\frac{y-Xw-\mu}{\sigma^2})\\
\mu n &= \sum y + w\sum X\\
\hat{\mu}_{MLE} &= \frac{\sum y + w\sum X}{n}
\end{aligned}
$$

Ans 3. See attached jupyter notebook

Ans 4.



The Network is an Adder. It has 2 XOR, 1 AND and 1 OR gate.

h1-h2-h3 represent XOR gate 1.

h4 represents an AND gate

h5 represents an OR gate

h6-h7-h8 represent XOR gate 2.

let $\sigma(x) = 1, x \geq 1$

$= 0x < 1$

$h1 = \sigma(2u + 2w - 1)$

$h2 = \sigma(-2u - 2w + 3)$

$h3 = \sigma(2h1 + 2h2 - 3)$

$h4 = \sigma(v + x - 1)$

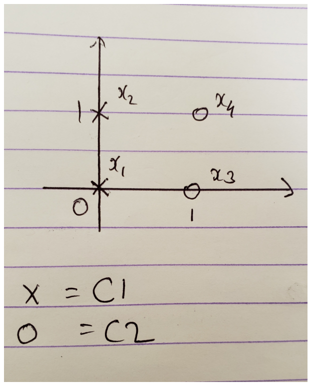
$h5 = \sigma(h3 + h4)$

$h6 = \sigma(2v + 2x - 1)$

$h7 = \sigma(-2v - 2x + 3)$

$h8 = \sigma(2h6 + 2h7 - 3)$

Ans 5. *a).* ☐



b).

Let C1 be denoted by -1 and C2 by 1

Steps Used:

1.Take dot product of W and X to find y

2.If dot product ≥ 0 then y=1 else y=-1

3.$W_{new} = W + \eta(d-y)X^T$

W	X	d	y	d-y
[0,0,0]	[1,0,0]	-1	-1	0
[0,0,0]	[1,0,1]	-1	-1	0
[0,0,0]	[1,1,0]	1	-1	2
[1,1,0]	[1,1,1]	1	1	0
[1,1,0]	[1,0,0]	-1	1	-2
[0,1,0]	[1,0,1]	-1	-1	0
[0,1,0]	[1,1,0]	1	1	0
[0,1,0]	[1,1,1]	1	1	0
[0,1,0]	[1,0,0]	-1	-1	0

Since, there is no change in the last 4 steps we stop

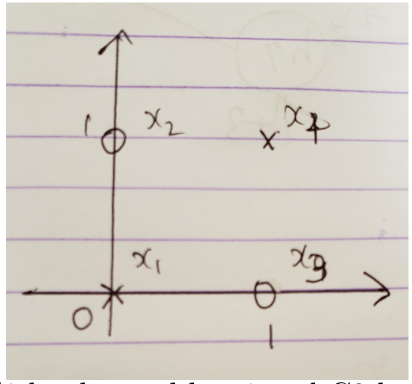
☐

c).

Decision Boundary is, $x = 0$

☐

d).



Let C1 be denoted by -1 and C2 by 1

W	X	d	y	d-y
[0,0,0]	[1,0,0]	1	-1	2
[1,0,0]	[1,0,1]	-1	1	-2
[0,0,-1]	[1,1,0]	-1	-1	0
[0,0,-1]	[1,1,1]	1	-1	2
[1,1,0]	[1,0,0]	1	1	0
[1,1,0]	[1,0,1]	-1	1	-2
[0,1,-1]	[1,1,0]	-1	1	-2
[-1,0,-1]	[1,1,1]	1	-1	2
[0,1,0]	[1,0,0]	1	-1	2
[1,1,0]	[1,0,1]	-1	1	-2
[0,1,-1]	[1,1,0]	-1	1	-2
[-1,0,-1]	[1,1,1]	1	-1	2
[0,1,0]	[1,0,0]	1	-1	2

As we can see that the weights are repeating we can say that Perceptron will not converge. Therefore, we cannot find a decision boundary for XOR

□